

# The effectiveness of robotic training depends on motor tasks characteristics

Laura Marchal-Crespo<sup>1,2,3\*</sup>, Nicole Rappo<sup>1</sup>, Robert Riener<sup>1,2</sup>

<sup>1</sup> Sensory-Motor Systems (SMS) Lab, Institute of Robotics and Intelligent Systems (IRIS), Department of Health Sciences and Technology (D-HEST), ETH Zurich, Switzerland

<sup>2</sup> Reharobotics Group, Spinal Cord Injury Center, Balgrist University Hospital, Medical Faculty, University of Zurich, Switzerland

<sup>3</sup> ARTORG Center for Biomedical Engineering Research, Medical Faculty, University of Bern, Switzerland

\* Correspondence:

[laura.marchal@hest.ethz.ch](mailto:laura.marchal@hest.ethz.ch)

**Keywords: motor learning, haptic guidance, error amplification, motivation, rehabilitation robotics, feedback motor control, feedforward motor control, discrete task, continuous task, task characteristics**

## Abstract

Previous research suggests that the effectiveness of robotic training depends on the motor task to be learned. However, it is still an open question which specific task's characteristics influence the efficacy of error-modulating training strategies. Motor tasks can be classified based on the time characteristics of the task, in particular the task's duration (discrete vs. continuous). Continuous tasks require movements without distinct beginning or end. Discrete tasks require fast movements that include well-defined postures at the beginning and the end.

We developed two games, one that requires a continuous movement –a tracking task– and one that requires discrete movements –a fast reaching task. We conducted an experiment with thirty healthy subjects to evaluate the effectiveness of three error-modulating training strategies –no guidance, error amplification (i.e. repulsive forces proportional to errors) and haptic guidance– on self-reported motivation and learning of the continuous and discrete games.

Training with error amplification resulted in better motor learning than haptic guidance, besides the fact that error amplification reduced subjects' interest/enjoyment and perceived competence during training. Only subjects trained with error amplification improved their performance after training the discrete game. In fact, subjects trained without guidance improved the performance in the continuous game significantly more than in the discrete game, probably because the continuous task required greater attentional levels. Error-amplifying training strategies have a great potential to provoke better motor learning in continuous and discrete tasks. However, their long-lasting negative effects on motivation might limit their applicability in intense neurorehabilitation programs.

## 1. Introduction

Motor learning –i.e. the acquisition of new motor skills– is a crucial aspect in several research fields, such as psychology, neurology, sport training and rehabilitation research (Muratori et al. 2013). These disciplines aim to better understand how to provoke long term changes in the brain structures (i.e. neuroplasticity) in order to improve or gain new motor and cognitive skills. In particular, motor learning theories have been used as an important tool to investigate, improve and adapt neurorehabilitation exercises (Dietz and Ward 2015). There is increasing interest in using robotic devices to train new motor skills, since robots can physically interact with subjects in order to assist or resist during training of complex movements, such as multi-joint arm movements (Nef et al. 2009; Klein et al. 2010) and walking (Riener et al. 2005; Rüdte et al. 2016). There has been a progression in the development of robotic training strategies that specify how these robotic devices should interact with subjects in order to enhance motor learning and rehabilitation outcomes –see Marchal-Crespo and Reinkensmeyer (2009) for a review. One way to group robotic training strategies is according to their use of the kinematic errors. Robotic haptic guidance is a training strategy in which a machine physically interacts with the subject's limbs to provide guidance in a desired movement by reducing ongoing errors. Error augmenting strategies, on the contrary, are in some ways the opposite of haptic controllers, because they make movement tasks more challenging.

It is thought that by moving patients' limbs in ways that they would otherwise not be able to move might motivate them to engage in repetitive and intensive practice (Reinkensmeyer and Housman 2007). However, robotic guidance also appears to decrease physical effort during training (Israel et al., 2006) suggesting that robotic training could potentially decrease learning if it encourages subjects' slacking (Reinkensmeyer et al., 2009; Scheidt et al., 2000). In fact, there is little evidence that haptic guidance is beneficial for human motor learning beyond enhancing safety and motivation. The long-standing hypothesis in motor learning research –the Guidance Hypothesis– states that physically guiding a movement will impair motor learning, because the user learns to rely on the guidance and fails to learn the motor commands required to perform the task (Salmoni et al. 1984; Schmidt and Bjork 1992).

Error amplification, on the other hand, is based on the idea that errors are needed in order to improve motor learning (Emken et al. 2007; Kitago and Krakauer 2013). Experimental evidence has demonstrated that it is possible to accelerate adaptation processes in healthy subjects when errors are accentuated using robotic forces during walking (Emken and Reinkensmeyer 2005). In patients with chronic stroke, amplifying errors during reaching with a force field resulted in straighter movements when the force field was removed (Patton et al. 2006). Training with error augmentation also benefited learning to play a pinball-like game in initially more skilled subjects (Milot et al. 2010). However, there are also studies that did not find a benefit from augmenting errors during training. Augmenting errors with forces proportional to errors did not enhance learning to track a 2D figure (Lee and Choi 2010). Amplifying errors during training how to synchronize the legs to track a Lissajous figure resulted in better learning in initially more skilled subjects, but limited transfer of learning (Marchal-Crespo et al. 2014a, 2017). These contradictory results suggest that error amplification might limit learning of some specific motor tasks, such as figure tracking. Augmenting kinematic errors during training seems to enhance the motor learning process by making the motor system to respond more strongly (Patton et al. 2006; Emken et al. 2007). However, in continuous

tasks where feedback is continuously provided, subjects might rely on the enhanced feedback and fail to learn the motor task.

Motor control is the process by which humans use their central nervous system (i.e. brain/cognition) to control the musculoskeletal system (i.e. muscles and limbs) involved in the performance of a motor skill. Motor learning is thought to be controlled by two different mechanisms in the brain: Feedback and feedforward motor control. Feedback motor control enables the motor system to respond to sensory information about deviations from the planned movement. Therefore, it depends on the sensory system to detect errors and perturbations (Seidler et al. 2004). As soon as errors can be detected through the sensory system, the information is used to adapt the motor commands for consecutive actions (Seidler et al. 2013). However, feedback control adds time delays into the motor response that might lead to inaccurate movement adjustments (Wolpert et al. 2001). The motor system is able to deal with these time delays employing feedforward motor control. Feedforward motor control works in a more predictive manner, i.e. it uses an internal model to predict the outcome of future movements (Kawato 1999; Tong and Flanagan 2003; Emken and Reinkensmeyer 2005). There is large evidence that the cerebellum (Wolpert et al. 1998; Shadmehr et al. 2010), parietal cortex, Basal ganglia and cingulate cortex are important brain regions involved in error-based learning –see Seidler et al. (2013) for a review. Each brain area has its specific characteristics while performing a task that requires feedback or feedforward motor control (Seidler et al. 2013). Therefore, it is unlikely that a training strategy that enhances learning of tasks requiring mostly feedforward control (e.g. discrete tasks, such as fast reaching or hitting tasks) will also benefit learning of tasks that mainly require feedback motor control (e.g. continuous tasks, such as tracking tasks). Therefore, the contradictory results observed in previous motor learning studies might originate from the fact that these experiments employed motor tasks that were governed by different motor control processes.

There is an initial body of work that compared the effectiveness of robotic strategies that augment or reduce movement errors on motor learning (Lee and Choi 2010; Sigrist et al. 2015). However, results from these comparative experiments are highly inconclusive, probably because they searched for the strategy that enhances learning, independently of the subjects' skill level and the specific characteristics of the task to be learned (Marchal-Crespo et al. 2015). Haptic guidance seems to be particularly helpful for initially less skilled subjects (Marchal-Crespo et al. 2010b, 2013), while error amplification was found to be more beneficial for more skilled participants (Milot et al. 2010; Duarte and Reinkensmeyer 2015). Additionally, error-augmenting strategies might be more suitable to enhance learning of especially simple tasks, i.e. tasks that can be mastered in a single practice session, since it might increase subjects' motivation and concentration (Marchal-Crespo et al. 2014b). On the other hand, in more challenging tasks, augmenting errors might decrease feelings of perceived competence and satisfaction and result in a long-term decrease in motivation that might limit the effectiveness of error amplification on motor learning (Duarte and Reinkensmeyer 2015).

To date, there is not a good understanding of the role of the motor tasks to be learned on the effectiveness of robotic training on motor learning. Thereby, current robotic devices might be working with suboptimal training strategies – only using a fraction of the potential – by not considering the specific characteristics of the task to be learned. The goal of this paper is to evaluate the impact of three error-modulating training strategies –no guidance, error amplification (i.e.

repulsive forces proportional to errors) and haptic guidance— on self-reported motivation and learning of two virtual reality (VR) games: A tracking game (continuous task) and a fast reaching game (discrete task). Both games required that the subjects move their dominant arm with only one degree of freedom (vertical or horizontal direction). In order to rule out the effect of the initial skill level on learning, we adapted the games difficulty to the subject's initial skill level. We conducted a motor learning experiment with thirty healthy subjects using ARMin IV, a robotic exoskeleton for upper limb rehabilitation. We hypothesized that after adapting the games difficulty to the subjects' initial skill levels, error amplification would enhance motor learning of the discrete tasks but would not have a positive effect on learning the continuous task. We also expected a decrease in self-reported feelings of enjoyment and perceived competence and an increase of effort during training with amplified errors. Finally, we hypothesized that the continuous task game would require greater attention levels, indicated by higher values of self-reported effort and pressure compared to the discrete task game.

## 2. Methods

### 2.1 ARMin IV

ARMin IV was employed to conduct the motor learning experiment (Fig. 1). ARMin IV is a 7 degree-of-freedom (DoF) robotic device for upper limb rehabilitation designed by the SMS-lab at ETH Zurich in collaboration with the University Hospital Balgrist (Nef et al. 2009). The robot incorporates position and force/torque sensors to measure the interaction between human and robot which enable subjects to perform a wide range of natural movements of the upper extremities (Guidali et al. 2011). A real time system with a sampling rate of 1 kHz controls the robotic device and communicates with the virtual world over a UDP connection.

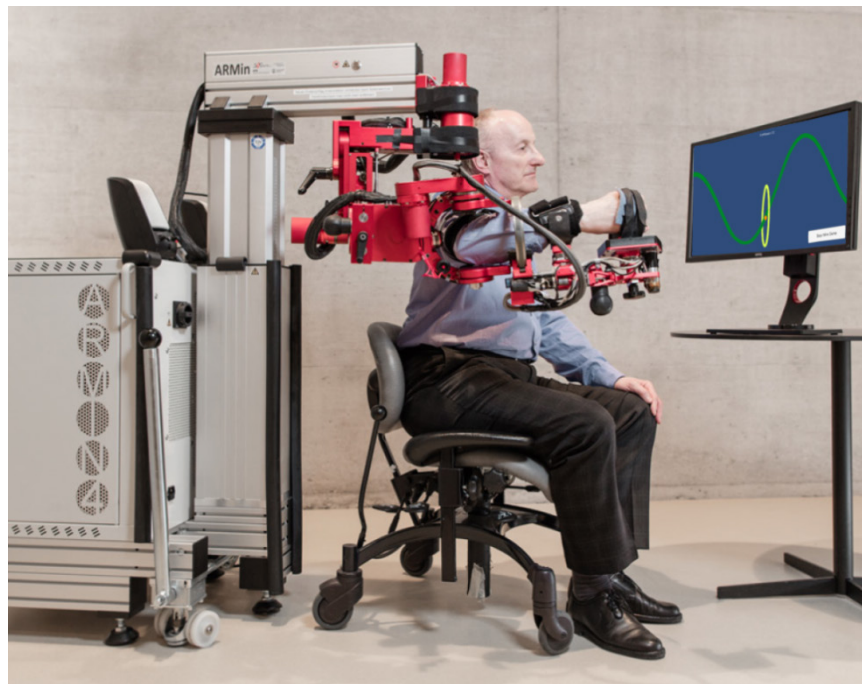


Fig. 1: ARMin IV is a 7 DoF robotic device designed for upper limb rehabilitation (Nef et al. 2009).

## 2.2 Continuous and discrete motor tasks

Two VR games were designed using Unity© 5.0 (Unity Technologies, United States): a continuous task game –the wire loop– and a discrete task game –the bar game. Both games are relatively similar to each other and do not employ complicated visual effects in order to reduce the effect of external factors (e.g. graphics attractiveness, motivational sounds) on the subjects' motivation and learning outcomes. In both games, the subject's right hand position is represented by a yellow ring with an orange dot at the center which can be moved in the vertical direction by moving the hand up and down (Fig. 2). Both games are represented in 2D on the computer screen, but the ring can only be moved in the vertical axis.

### 2.2.1 Continuous motor task game: the wire loop

A continuous task which is mainly controlled through feedback motor control should provide subjects with sensory information (e.g. the deviations between actual and desired position) during practice. This error signal can then be used by the motor system to appropriately adapt the ongoing movement. In addition, the task should not be predictable in order to require the subjects' full attention and prevent changing from feedback motor control to feedforward motor control as training progresses.

The goal of the wire loop game is to move the arm up and down in order to align the center of the ring (represented by an orange dot) on the green wire to avoid the ring to touch the wire (Fig. 2A). The wire moves from the right- to left-hand side of the screen at a speed that depends on the subject's initial skill level (see Section 2.2.3). The shape of the wire follows the addition of two sinusoids of different amplitude (12.5 cm and 5 cm) and frequency ( $2 \cdot f$  Hz and  $f$  Hz, with  $f$  depending on the wire speed). The resulting wire shape is not trivial, and therefore, it is unlikely that subjects can learn the shape, since they only see a small part of the wire at a time (i.e. the entire wire's trajectory is not visible during the game). The diameter of the ring is 18 cm.

While playing the wire loop game, subjects received continuous visual feedback about the actual position (center of the ring) and the desired position (green wire). In addition, subjects received information about the collisions: As soon as the ring touched the wire, the wire changed the color to red. The number of collisions were counted and displayed on the screen (Fig. 2A).

### 2.2.2 Discrete motor task game: the bar game

Discrete tasks require fast movements that include well-defined postures at the beginning and the end of the movement. We were especially interested in discrete tasks that are mainly controlled through feedforward control. Therefore, the discrete task should not provide subjects with feedback during task execution. The goal of the bar game is the same as in the wire loop game, move the arm up and down in order to align the ring center (orange dot) on the green wire and, therefore, avoid the ring to touch the wire (Fig. 2B). The wire is composed by a series of bars randomly located at three different vertical positions that move from the right- to left-hand side of the screen at a speed of 20 cm/s. Subjects were instructed to move the ring vertically in one run and as fast as possible to the next bar as soon as the ring reached the right end of the bar, minimizing the distance from the center of the ring to the next bar (Fig. 2B). While subjects moved the ring to the next bar, there was an additional

challenge: As soon as the ring leaved the right end of the bar, the next bar disappeared, and therefore, subjects performed the planned movement based on their memory and experience without visual feedback. The bar reappeared when the left end of the next bar reached the ring horizontal position, and therefore, subjects received terminal visual feedback about how precise their movement was once the movement was finished. The distance between the lowest and highest bar is 20 cm and between the lowest and middle bar is 15 cm. The bars length depends on the initial skill level of subjects (see Section 2.2.3).

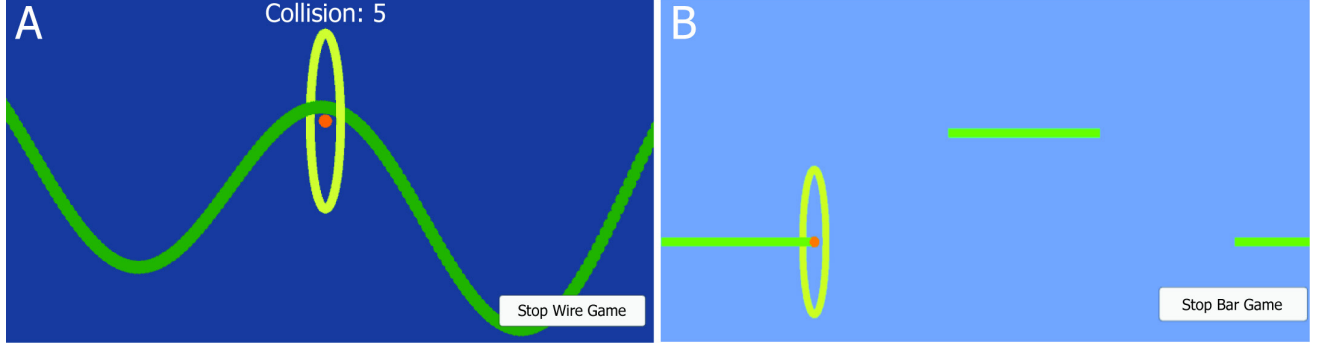


Fig. 2. **A:** The wire loop game. The goal is to move the arm up and down in order to center the ring (orange dot) on the green wire. The wire moves from the right- to left-hand side of the screen. **B:** The bar game. The goal is to move the ring vertically in one run and as fast as possible to the next bar as soon as the ring reaches the right end of the bar, minimizing the distance from the center of the ring (orange dot) to the next bar. As soon as the ring leaves the right end of a bar, the next bar disappears. The bar reappears when the left end of the next bar reaches the ring.

### 2.2.3 Adaptive adjustment of game difficulty

The difficulty levels of the continuous and discrete games were adjusted prior to training in order to control for the effect of game difficulty on learning. Adjustment was performed based on ongoing subject's performance during a first baseline test.

In the continuous (wire) game, the game difficulty was adapted by changing the speed of the wire, which moved from right to left on the screen, i.e. higher speeds increased the game difficulty. The subject's performance was measured on-line as the running mean absolute difference between the desired position (wire) and the actual position (ring center). The wire speed adjustment was performed at each time step (every 0.02 s) as follows:

$$v_{i+1} = \begin{cases} v_i + \delta_v, & \text{if } \frac{1}{i} \sum_{k=1}^i |e_{vk}| < e_{thres} \text{ and } v_i < v_{max} \\ v_i, & \text{otherwise} \end{cases} \quad (1)$$

Each subject started with an initial wire speed of 5 cm/s. At each sample time  $i$ , the running average absolute vertical error was calculated. If this average error was smaller than a threshold ( $e_{thres} = 2.5$  cm) and the speed  $v_i$  was smaller than the maximum wire speed ( $v_{max} = 17.5$  cm/s), the speed was increased by  $\delta_v$  (0.0045 cm/s). The parameters  $e_{thres}$ ,  $v_{max}$  and  $\delta_v$  were chosen based on pilot experiments in such a way that subjects were able to easily adapt to a smoothly updated wire speed.

In the discrete (bar) game, the difficulty level was adapted by changing the length of the bars, i.e., shorter bars increased the game difficulty. Per each bar, the maximum absolute difference between

the ring center and the bar vertical position was calculated. Then, when the ring leaved a bar, the average maximum error over all preceding bars was calculated. The bar length was adjusted every time the ring leaved a bar ( $j$ ) as follows:

$$l_{j+1} = \begin{cases} l_j - \delta_l, & \text{if } \frac{1}{j} \sum_{k=1}^j \max_i |e_{vi}| < l_{thres} \text{ and } l_j > l_{min} \\ l_j, & \text{otherwise} \end{cases} \quad (2)$$

Each subject started with an initial bar length of 40 cm. When the ring leaved a bar, the average maximum absolute error over all preceding bars was calculated. If this average maximum error was smaller than a threshold ( $l_{thres} = 2.5$  cm) and the current bar length ( $l_j$ ) was larger than the minimum length ( $l_{min} = 15$  cm), the bar length  $l_j$  was decreased by  $\delta_l$  (0.99995 cm). The minimum bar length of 15 cm was selected based on pilot experiments that showed that playing with shorter bars was no longer feasible. The parameters  $l_{thres}$  and  $\delta_v$  were selected based on pilot experiments.

#### 2.2.4 Generalization games

Generalization is a crucial aspect of motor learning. It is a type of transfer of learning that occurs from one task to another very similar task (Schmidt and Lee 2010). In order to evaluate generalization, subjects played both games turned 90 ° counterclockwise, i.e. they played on the horizontal axis, by moving the hand right/left, while the wire/bars moved from the top of the screen. All other parameters, like the goal of both games, speed and bar length, were equal to the skill-adjusted original games.

### 2.3 Training Strategies

#### 2.3.1 No Guidance

The developed games required the subjects to control the ring position by moving the hand in the vertical direction. In order to avoid subjects to move in other directions but the vertical, we developed an impedance controller in end-effector coordinates with zero impedance gain in the vertical direction, and high stiffness gains in the horizontal plane. The impedance forces were then transformed into joint torques by using inverse kinematics and input to a close-loop torque controller. Ideally, the robot should be transparent when the impedance is zero, i.e. the interaction torques between robot and patient should be zero. However, due to the mass and friction of the exoskeleton, transparency cannot be fully achieved through impedance control. Therefore, the force of gravity acting on the exoskeleton (Nef et al., 2009) and the forces due to friction of the motor-gear combinations were modeled and used as feedforward compensation (Guidali et al., 2011).

#### 2.3.2 Haptic Guidance

The goal of the haptic guidance strategy was to slightly guide subjects in accomplishing both feedback and feedforward games, i.e. reduce the errors, rather than remove them. Therefore, an impedance controller with a relatively small impedance gain in the vertical direction ( $k_H = 4$  N/cm), and high stiffness gains in the horizontal plane was implemented to assists subjects as needed. The forces exerted by the vertical haptic guidance controller ( $F_H$ ) can be described as:

$$F_H = k_H \cdot (x_d - x) = k_H \cdot e_v, \quad \text{where : } k_H > 0. \quad (3)$$

Equation (3) describes a position controller in end-effector coordinates where the error in vertical direction ( $e_v$ ) is reduced with a positive proportional gain  $k_H$ . In the feedback game, the desired position ( $x_d$ ) was defined by the position of the wire at the current time step. The vertical error ( $e_v$ ) was calculated as the difference between the position of the ring center ( $x$ ) and the wire ( $x_d$ ). In the feedforward game, the desired position ( $x_d$ ) was defined by the bar position only when the bar was visible. During the fast movements between two consecutive bars, the desired trajectory was calculated following the minimum jerk equation, i.e., a straight vertical movement between the vertical positions of the previous bar and the consecutive (invisible) bar with a bell shape velocity profile (Flash and Hogan 1985). The mean target movement speed was chosen ad hoc to be 24 cm/s.

The haptic force in end-effector coordinates was saturated at 10 N to guarantee subjects' safety, transformed into joint torques using inverse kinematics and input to a close-loop torque controller, where friction and weight compensation were added (see section 2.3.1).

### 2.3.3 Error Amplification

In order to augment errors, a proportional controller with negative impedance gain in the vertical direction ( $k_{EA} < 0$ ) was developed. This controller worked similarly to the haptic guidance one, but the vertical forces directed the ring away from the wire. The forces exerted by the error amplification controller ( $F_{EA}$ ) can be described as:

$$F_{EA} = k_{EA} \cdot (x_d - x) = k_{EA} \cdot e_v, \quad \text{where : } k_{EA} < 0. \quad (4)$$

The desired ( $x_d$ ) and actual positions ( $x$ ) and vertical error ( $e_v$ ) were calculated as described in previous subsection 2.3.2. The error amplification forces increase proportionally to vertical errors. Since subjects have a maximum force that they can apply, the system could potentially make the ring leave the wire. In order to limit large errors that can be too dangerous or discouraging for subjects, error amplification should be applied only in a small predefined area around the wire. By additionally realizing a conversion towards haptic guidance when the error is larger than the ring radius, the system could prevent the ring to leave the wire. Therefore, we made the proportional gain  $k_{EA}$  a function of the vertical error (Rüdt et al. 2016). The used equation corresponds to the superposition of two sigmoid functions (Fig. 4):

$$k_{EA} = K_{\max} \cdot \left( \frac{2}{1 + \exp(r \cdot (|e_v| - R))} - 1 \right). \quad (5)$$

The resulting function can be tuned by three parameters. The parameter  $K_{\max}$  stands for the maximal amplification gain when the vertical error  $e_v$  is equal to zero. The parameter  $R$  defines the error where the gain  $k_{EA}$  changes its sign (ring radius). Due to this change, the amplification force also changes its sign, transitioning from error amplification (EA) to haptic guidance (HG) (Fig. 3 bottom). The combined sigmoid function has saturation for small and large  $e_v$ . The width of the saturated area and the steepness of the curve around  $R$  can be tuned using the parameter  $r$ .

The maximal amplification gain  $K_{\max}$  was selected ad hoc. The error amplification controller should allow subjects to make small errors in order to stimulate motor learning. Nevertheless, subjects should



also be able to stay on the desired target position. In order to enable a certain tracking accuracy without an excessive physical effort from the subject, a small  $K_{max}$  gain of 1 N/cm was selected. In order to provide a good compromise between smooth transition and unmodified control modes, the slope  $r$  of the sigmoid was set to 2 (Fig. 3 top).

The error amplification force in end-effector coordinates was saturated at 10 N to guarantee subjects' safety, transformed into joint torques using inverse kinematics and input to a close-loop torque controller (section 2.3.1).

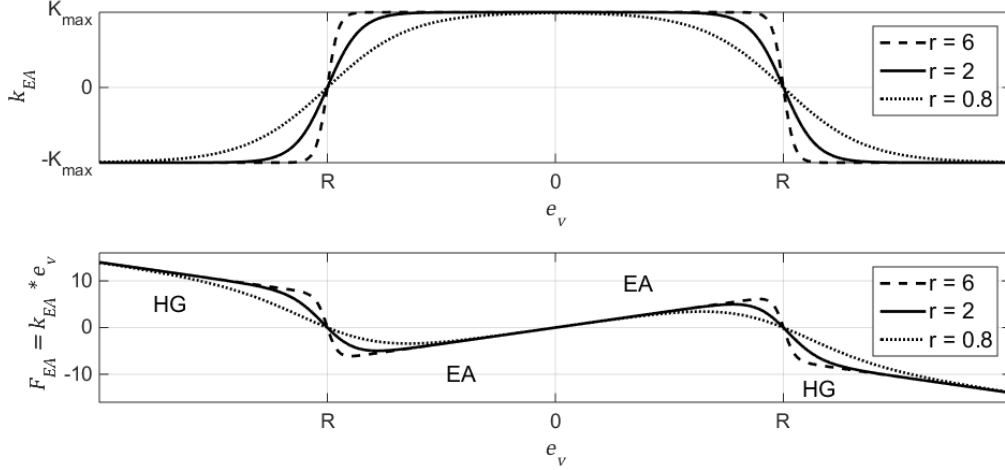


Fig. 3: Combined sigmoid function for the gain  $k_{EA}$  (top) and its resulting vertical force  $F_{EA}$  (bottom). The proportional gain  $k_{EA}$  is a function of the subject's ongoing vertical error  $e_v$ . It prevents errors bigger than the ring radius using haptic guidance (HG), while augments awareness of smaller task-relevant errors by means of error amplification (EA).

## 2.4 Experimental Protocol

The study was approved by the local ethical committee (Kantonale Ethikkommission Zürich, KEK-ZH-Nr. 2015-0013) and conducted in compliance with the Declaration of Helsinki. Thirty healthy, right-handed adult subjects without any history of neurological disease (22 men),  $26.7 \pm 2.8$  y.o., gave written consent to participate in the study. Subjects were randomly assigned to one of three training strategies: No guidance (NG), haptic guidance (HG), and error amplification (EA).

Subjects sat in front of a computer screen. The subjects' right arm was positioned in ARMin IV with an initial elbow excursion angle of  $-10^\circ$  and shoulder elevation  $75^\circ$  (Wu et al. 2002). They were informed that during practice the robot could either help or disturb them and to be as active as possible. In the feedback game, subjects were instructed to move their hand on the vertical axis to constantly locate the center of the ring on the green wire. They were informed that the wire color would change to red if they collide with the wire and that the number of collisions would be counted. In the feedforward game, subjects were instructed to move the ring as fast as possible in one run to the next bar as soon as they reached the right end of a bar. They were informed that the next bar would disappear once the ring leaves the right end of the previous bar and it would reappear when the ring reaches the new bar. In order to prevent the game order to influence on the results, subjects in each group were randomly assigned into two subgroups, which practiced the games in contrary order.

The experiment took place in two different days. The protocol (Fig. 4) was performed separately for each game. On day 1, subjects started with a baseline test (1 min, B1), in which the difficulty level of the game was adapted to their initial skill level (see section 2.2.3). Once the wire speed –or bar length– was established, a baseline test with the adapted value was performed (1 min, B2). After a short break, subjects performed the baseline generalization test (1 min, BG), where they played again the same game but 90 ° rotated counterclockwise. During all baseline tests, the robot worked with the NG controller. During training, subjects played with NG, HG or EA, depending on their training strategy group, in four blocks of 1 min each (T), and 30 s breaks between them. After a five minute break, short-term retention (ST) and short-term retention generalization (STG) tests were performed that followed the same structure and order than B2 and BG tests. The same protocol for the second game was then performed. The total experimental time was around 1 hour. Subjects were invited to return after 3-5 days to perform the retention (RT) and generalization retention (GRT) tests, which followed the same structure and order as ST and STG tests.

After the second baseline (B2), after the last training block (T4) and after the retention test (RT) in each game, subjects responded to 11 questions (Table 1) selected from the Intrinsic Motivation Inventory (IMI, Ryan 1982), which is a well established and valid questionnaire already used in previous motor learning and rehabilitation studies (Novak et al. 2014; Duarte and Reinkensmeyer 2015). The full questionnaire consists of 45 questions and can be further divided into seven subscales. In the present study, only four subscales were selected: Interest/enjoyment (I/E), perceived competence (PC), effort/importance (E/I) and pressure/tension (P/T). These four items were selected because they encompass the essentials of the motivational aspects in the specific motor task (McAuley et al. 1989). Subjects rated their feelings on a Likert scale between 1 and 7 points; 1 indicated “I disagree completely” and 7 indicated “I agree completely”. The questions were presented in German and English. Subjects could see their previous answers to the same questions.

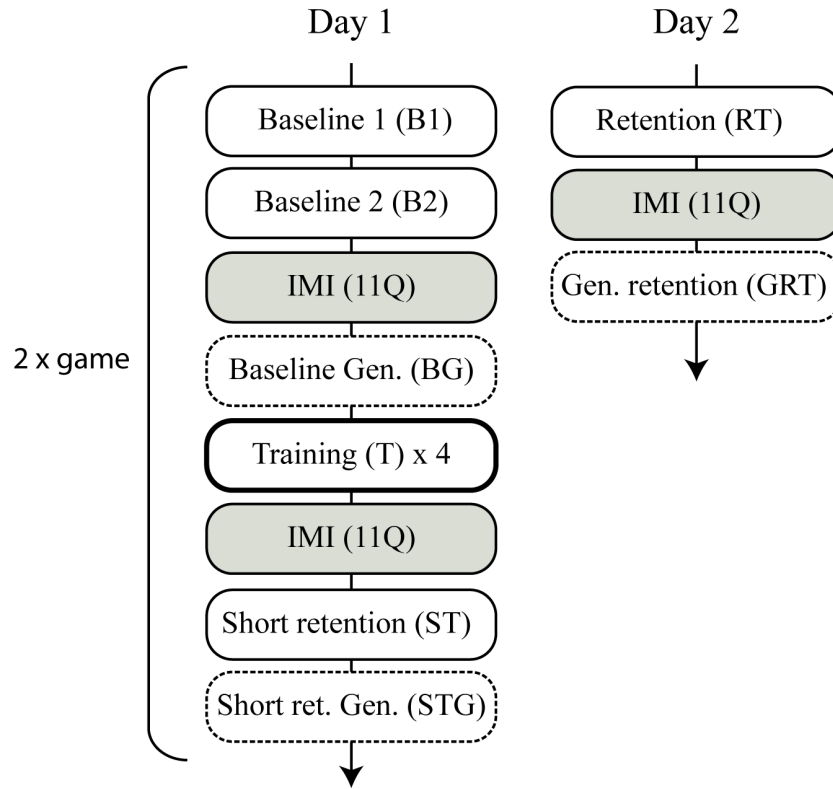


Fig. 4: Experimental protocol. On day 1, subjects started with a baseline test (1 min, B1), in which the difficulty level of the game was adapted to their initial skill level. A baseline test with the adapted game parameters was then performed (1 min, B2). After a short break, subjects performed the baseline generalization test (1 min, BG). During training, subjects trained with NG, HG or EA in four blocks of 1 min each. After a 5 min break, short-term retention (ST) and short-term retention generalization (STG) tests were performed. The same protocol for the second game was then performed. Subjects returned after 3-5 days to do the retention tests (RT and GRT) for both games. After B2, after the last training block (T4) and after RT in each game, subjects responded to 11 IMI questions.

Table 1. IMI Questionnaire

IMI Questions	Subscale
1. I enjoyed playing the game very much	interest/enjoyment (I/E)
2. I am satisfied with my performance in the game	perceived competence (PC)
3. I put a lot of effort into this	effort/importance (E/I)
4. I think the game was boring	(I/E)
5. I felt very tense while playing the game	pressure/tension (P/T)
6. I could not play the game very well	(PC)
7. It was important for me to do well at the game	(E/I)
8. This game was fun to play	(I/E)
9. I was anxious while playing the game	(P/T)
10. I think I am pretty good at the game	(PC)
11. I did not try very hard to do well at the game	(E/I)

Table 1: Set of 11 questions selected from the Intrinsic Motivation Inventory (IMI, Ryan 1982) in order to assess subjects' interest/enjoyment (I/E), perceived competence (PC), effort/importance (E/I) and pressure/tension (P/T).

## 2.5 Data Processing and Statistical Analysis

The subjects' performance at each protocol block (Fig. 4) in the continuous (wire) game was calculated as the mean absolute difference between the desired position (wire) and the actual ring position. In the discrete (bar) game, the performance variable was calculated as the mean absolute difference between the desired position (green bar) and the ring position at the moment when the ring reaches the left end of the bar –i.e. when the target bar reappears. Therefore, the error is calculated before subjects could correct their movements using visual feedback. For both games, the standard deviation of the signed errors (i.e. error variability) was calculated. The error variability is an important indicator of learning, since it is a measure of how consistent the subjects' movements were.

We used linear mixed models to test the effect that the three training strategies (no guidance, haptic guidance and error amplification) and the game (continuous and discrete) had on the change in the mean error and variability between baseline 2 (B2) and the first training block (T1), short-term (ST) and retention tests (RT) and between baseline generalization (BG) and short-term (STG) and generalization retention tests (GRT) (therefore, a total of 5 LMM for each performance variable). Linear mixed models provide asymptotically normal estimates and a consistent covariance matrix so that simultaneous inference procedures can be set up (Hothorn et al. 2008). In order to answer our specific hypothesis (listed in Table 2), we created a new categorical super-factor (“TrainingGameFactor”) with six different levels corresponding to each of the different combinations between the game and the corresponding training strategy (Table 2). We fitted our data ( $y_{ij}$ ) using a linear mixed model of the form:

$$y_{ij} = \mu + \beta_i + \gamma_j + \epsilon_{ij} \quad (6)$$

where  $\mu$  is the global mean,  $\beta_i$  (with  $i = 1, \dots, 6$ ) is the super-factor fixed effect (“TrainingGameFactor”),  $\gamma_j$  (with  $j = 1, \dots, 30$ ) is the subject random effect with assumed normal distribution  $N(0, \sigma_\gamma^2)$ , and  $\epsilon_{ij}$  is the error term with assumed normal distribution  $N(0, \sigma_\epsilon^2)$ . Every person had two treatments (they played the two games) in random order. Only random by-subject adjustment of the intercept was considered whereas possible random by-subjects slopes adjustments were left out in the LMM models. We used the function `lmer` from the `lme4` package for R described in (Bates 2015) to fit our model (eq. 6). We performed post-hoc analysis in order to test a battery of 14 linear hypotheses (Table 2) using the General Linear Hypotheses function (`glht`) in R (Hothorn et al. 2008). The significant levels ( $p$ -values) were computed using z-test and adjusted using single step method. The matrix employed in the `glht` function is summarized in Table 2. In the results section, we report the estimates ( $\beta$ ), standard errors ( $SE$ ), z-values, and significant levels ( $p$ ).

Linear hypotheses	Explanation	TrainingGameFactor					
		Wire EA	Wire HG	Wire NG	Bar EA	Bar HG	Bar NG
Test 1	Compare EA vs. HG	1/2	-1/2	0	1/2	-1/2	0
Test 2	Compare EA vs. NG	1/2	0	-1/2	1/2	0	-1/2
Test 3	Compare Bar vs. Wire	-1/3	-1/3	-1/3	1/3	1/3	1/3
Test 4	Wire game: Compare EA vs. HG	1/2	-1/2	0	0	0	0
Test 5	Wire game: Compare EA vs. NG	1/2	0	-1/2	0	0	0
Test 6	Wire game: compare HG vs. NG	0	1/2	-1/2	0	0	0
Test 7	Bar game: Compare EA vs. HG	0	0	0	1/2	-1/2	0
Test 8	Bar game: Compare EA vs. NG	0	0	0	1/2	0	-1/2
Test 9	Bar game: Compare HG vs. NG	0	0	0	0	1/2	-1/2
Test 10	EA: Compare Wire vs. Bar game	1/2	0	0	-1/2	0	0
Test 11	HG: Compare Wire vs. Bar game	0	1/2	0	0	-1/2	0
Test 12	NG: Compare Wire vs. Bar game	0	0	1/2	0	0	-1/2
Test 13	Compare difference EA-HG between Bar and Wire	-1/2	1/2	0	1/2	-1/2	0
Test 14	Compare difference EA-NG between Bar and Wire	-1/2	0	1/2	1/2	0	-1/2

Table 2: Linear hypotheses and corresponding matrix employed as an input to the General Linear Hypotheses function (glht) in R. For each game, continuous (Wire) and discrete (Bar), three different training strategies were tested: Error amplification (EA), haptic guidance (HG) and no guidance (NG). A total of six different combinations (between the game and the corresponding training) are represented at the top-right of the table (TrainingGameFactor). Overall, 14 different hypotheses were tested.

We used ANOVAs to test whether the performance variables in each game during baseline 2 were different between training groups. Paired t-tests between baseline 2 and retention tests were performed to test if learning occurred after training with each strategy. We performed Pearson's correlation tests in order to evaluate the correlation between the mean absolute error during baseline 1 (B1) and the adapted speed (wire game) and bar length (bar game) reached at the end of B1.

We evaluated the effects that the different training strategies and games had on the four IMI subscales: interest/enjoyment (Q1, Q4, Q8), perceived competence (Q2, Q6, Q10), effort/importance (Q3, Q7, Q11) and pressure/tension (Q5, Q9). We compared the responses after baseline using nonparametric independent samples Kruskal-Wallis tests with training strategy as the main factor. We used the responses after baseline as reference point. We compared the changes in responses with respect to baseline using Kruskal-Wallis test with training strategy as the main factor. Mann-Whitney Test range test was used on pairwise comparisons if the Kruskal-Wallis was significant. We tested the effect of the game in the response changes using nonparametric related samples Wilcoxon signed rank tests.

The normality and homoscedasticity assumption of the residuals was visually inspected using Q-Q plots and residuals versus fits plots in order to get a general idea of the data distribution. Post hoc comparisons were performed with Tukey correction. Except for the battery of linear hypotheses tests, the alternative hypothesis was two-sided. The significance level was set to  $\alpha = 0.05$ . Statistical analysis was performed in R (version 3.3.0) and IBM® SPSS® Software (version 21, Chicago, IL).

### 3. Results

#### 3.1 Game difficulty adjustment

We found a significant correlation between the speed reached by the end of baseline 1 (B1) and the mean absolute error in B1 in the continuous game (wire) (Fig. 5A,  $r = -0.801$ ,  $p < 0.001$ ), and between the bar length at the end of B1 and the mean absolute error during B1 in the discrete game (bar) (Fig. 5B,  $r = 0.782$ ,  $p < 0.001$ ). Therefore, there was a significant correlation between initial skill level (i.e. error in B1) and wire speed and bar length after adaptation –the more skilled the subjects, the faster the wire moved and shorter the bars were.

In order to test whether subjects reached similar initial skill level after the adaptation of the speed and bar length, we assigned subjects into two groups based on their mean error during B1. We used the median of the error created by all subjects during B1 as the cut off value –i.e. the 15 subjects with errors smaller than the median were designated as skilled subjects, and the other 15 as non- skilled subjects.. We run t-tests to compare the mean error between skilled and non-skilled subjects, and found that the performance was significantly different during B1 (wire game:  $t(22.58) = -6.35$ ,  $p < 0.001$ ; bar game:  $t(21.93) = -6.55$ ,  $p < 0.001$ ), but significant differences were not visible in B2 (wire:  $t(27.18) = -0.38$ ,  $p = 0.71$ ; bar:  $t(27.88) = -0.91$ ,  $p = 0.372$ ), suggesting that skilled and non-skilled subjects performed similarly during B2. We did not find significant differences between training groups in the mean errors performed during B2 (ANOVA,  $p > 0.05$ ).

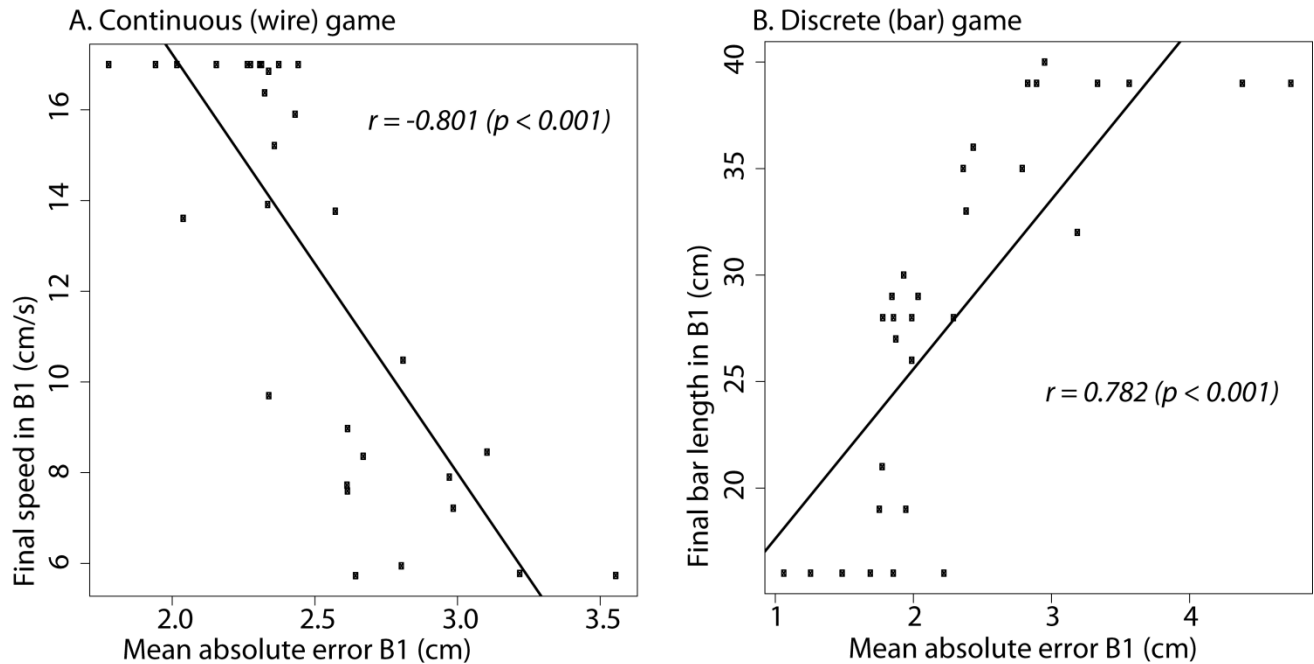


Fig. 5: Relation between adapted wire speeds (A) and bar lengths (B) at the end of B1 and mean absolute error during B1.

#### 3.2 Performance during training

When comparing the error changes from baseline 2 (B2) to the first training session (T1), a significant increase in the mean absolute error in the error amplification (EA) group (Fig. 6A & B, wire game:  $t(9) = -9.51$ ,  $p < 0.001$ ; bar game:  $t(9) = -3.88$ ,  $p = 0.004$ ) and a significant error

decrease in the haptic guidance (HG) group (Fig. 6 A & B, wire:  $t(9) = 8.53, p < 0.001$ ; bar:  $t(9) = 6.67, p < 0.001$ ) were observed. The no guidance group (NG) also reduced significantly the mean error in the wire game (Fig. 6A,  $t(9) = 3.48, p = 0.007$ ), but the error reduction was non-significant in the bar game. In fact, the error change from B2 to T1 was significantly different between training groups (wire:  $F(2,27) = 98.43, p < 0.001$ ; bar:  $F(2,27) = 29.39, p < 0.001$ ). Post hoc tests revealed that all pairwise comparisons between training groups were significantly different.

Similar results were found in the change of error variability: A significant increase in the EA group (Fig. 6 C & D, wire:  $t(9) = -5.42, p < 0.001$ ; bar:  $t(9) = -3.38, p = 0.008$ ) and a significant variability decrease in the HG group (Fig. 6 C & D, wire:  $t(9) = 11.57, p < 0.001$ ; bar:  $t(9) = 7.59, p < 0.001$ ). The NG group also reduced significantly the error variability in the wire game (Fig. 6C,  $t(9) = 3.22, p = 0.01$ ), but the change was non-significant in the bar game.

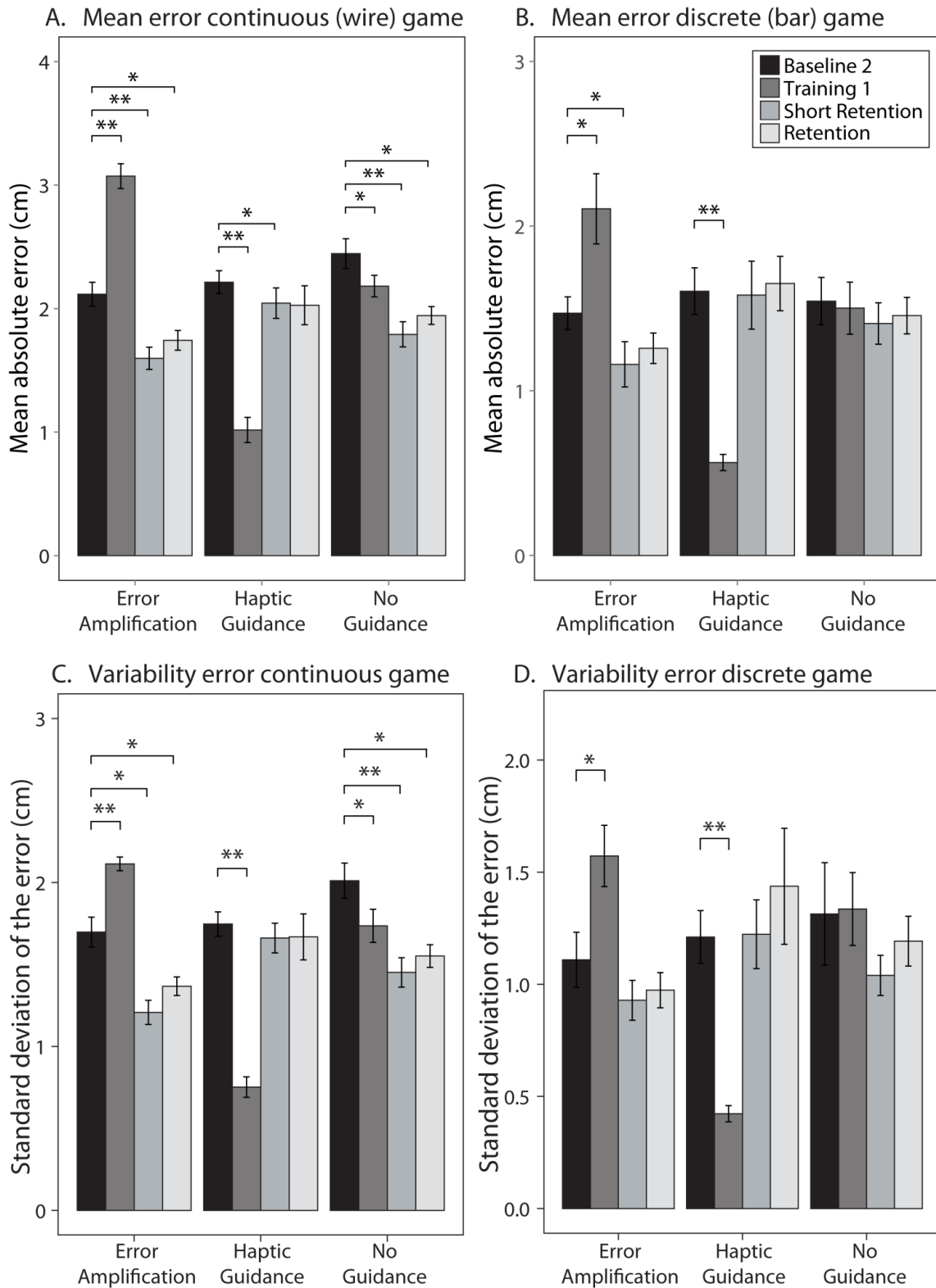


Fig. 6: Mean absolute error (A & B) and mean variability error (C & D) during baseline 2, first training block, and short-term and retention tests for subjects trained with error amplification, haptic guidance and no guidance, during the continuous (wire) game (A & C) and the discrete (bar) game (B & D). Error bars:  $\pm 1$  SE. \*\*  $p < 0.001$ , \*  $p < 0.05$ .



### 3.3 Effect of the training strategies and games on skill learning

In general, subjects learned the **continuous (wire) task**, as suggested by a significant mean error reduction from baseline 2 (B2) to short-term (ST) and - retention (RT) tests. In particular, all training groups reduced the mean errors significantly at ST (Fig. 6A, EA:  $t(9) = 5.02, p < 0.001$ ; HG:  $t(9) = 2.43, p = 0.038$ ; NG:  $t(9) = 4.86, p < 0.001$ ). However, at long term retention, only the EA (Fig. 6A,  $t(9) = 4.05, p = 0.003$ ) and NG groups (Fig. 6A,  $t(9) = 3.72, p = 0.005$ ) reduced the errors. In the error variability, however, only the EA and NG groups showed a significant decrease at ST (Fig. 6C, EA:  $t(9) = 4.15, p = 0.002$ ; NG:  $t(9) = 5.65, p < 0.001$ ) and RT (Fig. 6C, EA:  $t(9) = 3.11, p = 0.012$ , NG:  $t(9) = 4.54, p = 0.001$ ).

In the **discrete (bar) game**, only subjects trained with EA learned the task, i.e., they significantly reduced the mean error at ST (Fig. 6B,  $t(9) = 2.86, p = 0.018$ ). However, the mean error reduction at RT in the EA group did not reached significance (Fig. 6B,  $t(9) = 1.61, p = 0.142$ ). None of the training groups showed a significant decrease of the error variability (Fig. 6D, paired t-tests,  $p > 0.05$ ).

We found that subjects trained with EA reduced the mean error significantly more at ST than subjects trained with HG (Table 3 test 1,  $\beta = 0.319, SE = 0.120, z = 2.661, p = 0.042$ ). However, no significant differences were found between the EA and NG groups. We found a significant effect of the game on the error reduction at ST. Subjects reduced the mean error in a greater amount in the wire game compared to the bar game (Table 3 test 3,  $\beta = 0.291, SE = 0.088, z = 3.291, p = 0.006$ ). Interestingly, we found that only subjects in the NG group showed significant differences between the wire and bar games: They reduced the errors in the wire game in a greater amount than in the bar game (Table 3 test 12,  $\beta = 0.517, SE = 0.153, z = 3.380, p = 0.004$ ).

When evaluating the effect of the training strategies on the performance change at ST for each game independently (Table 3 tests 4-9), we found that the differences between EA and HG were more evident in the wire game (Table 3 test 4, mean error:  $\beta = 0.350, SE = 0.161, z = 2.168, p = 0.139$ ; error variability:  $\beta = 0.405, SE = 0.171, z = 2.374, p = 0.087$ ), although differences did not reached significance. We also found that in the wire game, subjects trained with HG reduced the mean error in a significant less amount than subjects trained with NG (Table 3 test 6,  $\beta = 0.484, SE = 0.161, z = 2.998, p = 0.015$ ), same with the error variability ( $\beta = 0.474, SE = 0.171, z = 2.779, p = 0.029$ ).

Most of the significant differences found between training strategies at short-term retention tests faded after 3-5 days at retention. We found a significant effect of game type on the error reduction at RT (Table 3 test 3,  $\beta = 0.269, SE = 0.102, z = 2.643, p = 0.044$ ).

Table 3.		B2-ST		B2-RT	
Linear hypotheses	Explanation	<i>p</i> -value mean error	<i>p</i> -value variability	<i>p</i> -value mean error	<i>p</i> -value variability
Test 1	Compare EA vs. HG	0.042 *	0.082 ·	0.288	0.273
Test 2	Compare EA vs. NG	0.989	1.000	0.997	1.000
Test 3	Compare Bar vs. Wire	0.006 *	0.077 ·	0.044 *	0.145
Test 4	Wire game: Compare EA vs. HG	0.139	0.087 ·	0.714	0.696
Test 5	Wire game: Compare EA vs. NG	1.000	1.000	1.000	1.000
Test 6	Wire game: compare HG vs. NG	0.015 *	0.029 *	0.296	0.363
Test 7	Bar game: Compare EA vs. HG	0.294	0.674	0.463	0.412
Test 8	Bar game: Compare EA vs. NG	0.700	1.000	0.881	0.996
Test 9	Bar game: Compare HG vs. NG	0.882	0.346	0.857	0.446
Test 10	EA: Compare Wire vs. Bar game	0.530	0.246	0.792	0.814
Test 11	HG: Compare Wire vs. Bar game	0.769	0.918	0.553	0.538
Test 12	NG: Compare Wire vs. Bar game	0.004 *	0.319	0.095 ·	0.446
Test 13	Compare difference EA-HG between Bar and Wire	0.999	1.000	0.978	0.972
Test 14	Compare difference EA-NG between Bar and Wire	0.492	0.999	0.743	0.953

Table 3: Linear hypotheses tests results from mean error and errors variability changes from baseline 2 (B2) to short-term (ST), and retention (RT) tests. Wire states for continuous game, Bar for discrete game. EA, HG and NG denote the training strategies: Error amplification, haptic guidance and no guidance. \*  $p < 0.05$ , ·  $p < 0.1$ .

### 3.4 Effect of the training strategies on generalization

In general, subjects generalized learning in the **Continuous (wire) game**, as suggested by a significant mean error reduction from baseline generalization (BG) and generalization short-term (STG:  $t(29) = 4.272$ ,  $p < 0.001$  and retention tests (GRT:  $t(29) = 4.218$ ,  $p < 0.001$ ). In particular, all training groups reduced the mean errors significantly at STG (Fig. 7A, EA:  $t(9) = 4.349$ ,  $p = 0.002$ ; HG:  $t(9) = 2.335$ ,  $p = 0.044$ ; NG:  $t(9) = 3.380$ ,  $p = 0.008$ ), and GRT (Fig. 7A, EA:  $t(9) = 5.488$ ,  $p < 0.001$ ; HG:  $t(9) = 4.681$ ,  $p = 0.001$ ; NG:  $t(9) = 2.891$ ,  $p = 0.018$ ). For the error variability, however, only the EA and the NG groups showed a significant decrease at STG (Fig. 7C, EA:  $t(9) = 6.291$ ,  $p < 0.001$ ; NG:  $t(9) = 2.422$ ,  $p = 0.038$ ). At GRT, only subjects in the EA and HG groups reduced significantly the error variability (Fig. 7C, EA:  $t(9) = 5.449$ ,  $p < 0.001$ ; HG:  $t(9) = 4.277$ ,  $p = 0.002$ ).

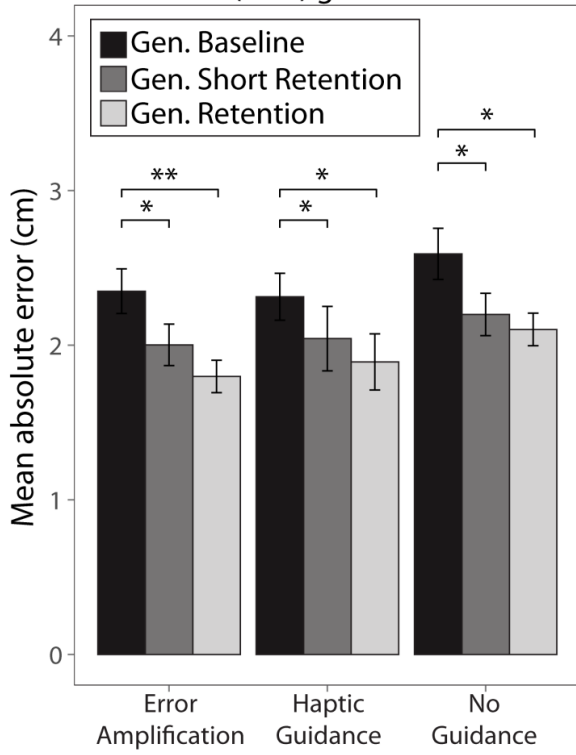
In the **discrete (bar) game**, only subjects trained with HG reduced significantly the mean error (Fig. 7B,  $t(9) = 3.485$ ,  $p = 0.007$ ) and error variability (Fig. 7D,  $t(9) = 2.946$ ,  $p = 0.016$ ) at GRT.

We did not find significant differences in any of the 14 linear hypotheses tested between BG-STG and BG-GRT, for the mean error reduction (table 4). For the error variability reduction, only the difference between HG and NG in the bar game at generalization retention reached significance (Table 4 test 9,  $\beta = 0.756$ ,  $SE = 0.246$ ,  $p = 0.012$ ).

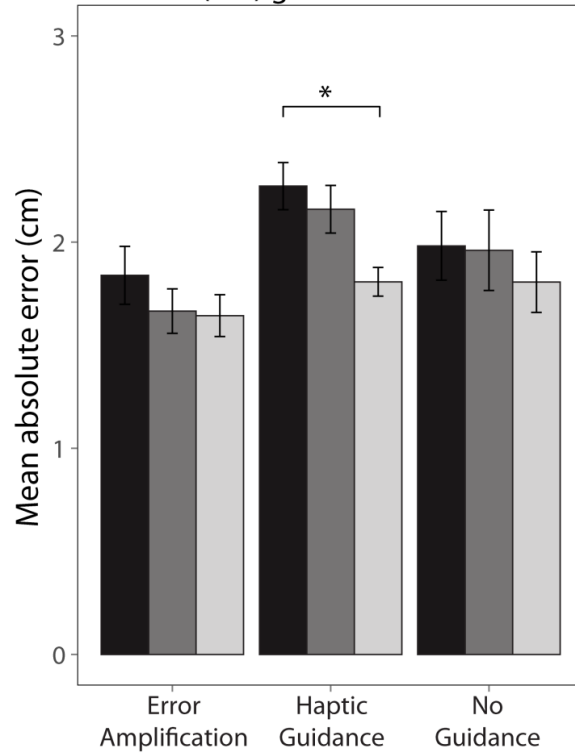
Table 4.		BG-STG		BG-GRT	
Linear hypotheses	Explanation	<i>p</i> -value mean error	<i>p</i> -value variability	<i>p</i> -value mean error	<i>p</i> -value variability
Test 1	Compare EA vs. HG	0.948	1.000	1.000	0.816
Test 2	Compare EA vs. NG	0.968	0.998	0.980	1.000
Test 3	Compare Bar vs. Wire	0.185	0.549	0.314	1.000
Test 4	Wire game: Compare EA vs. HG	0.967	0.876	0.910	1.000
Test 5	Wire game: Compare EA vs. NG	1.000	1.000	0.978	1.000
Test 6	Wire game: compare HG vs. NG	0.916	0.516	0.975	1.000
Test 7	Bar game: Compare EA vs. HG	0.977	1.000	1.000	0.189
Test 8	Bar game: Compare EA vs. NG	0.862	0.898	0.994	1.000
Test 9	Bar game: Compare HG vs. NG	1.000	1.000	1.000	0.012*
Test 10	EA: Compare Wire vs. Bar game	0.813	0.786	0.329	1.000
Test 11	HG: Compare Wire vs. Bar game	0.846	1.000	1.000	0.589
Test 12	NG: Compare Wire vs. Bar game	0.257	0.117	0.443	1.000
Test 13	Compare difference EA-HG between Bar and Wire	0.999	1.000	1.000	0.192
Test 14	Compare difference EA-NG between Bar and Wire	0.884	0.780	1.000	1.000

Table 4: Linear hypotheses tests results from mean error and errors variability changes from baseline generalization (BG) to short-term generalization (STG), and generalization retention (GRT) tests. Wire states for continuous game, Bar for discrete game. EA, HG and NG denote the training strategies: Error amplification, haptic guidance and no guidance. \*  $p < 0.05$

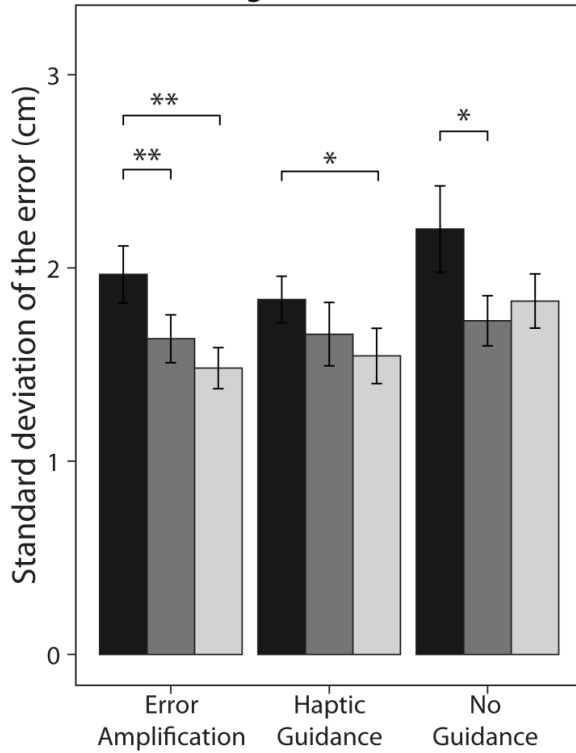
A. Generalization mean error continuous (wire) game



B. Generalization mean error discrete (bar) game



C. Generalization variability error continuous game



D. Generalization variability error discrete game

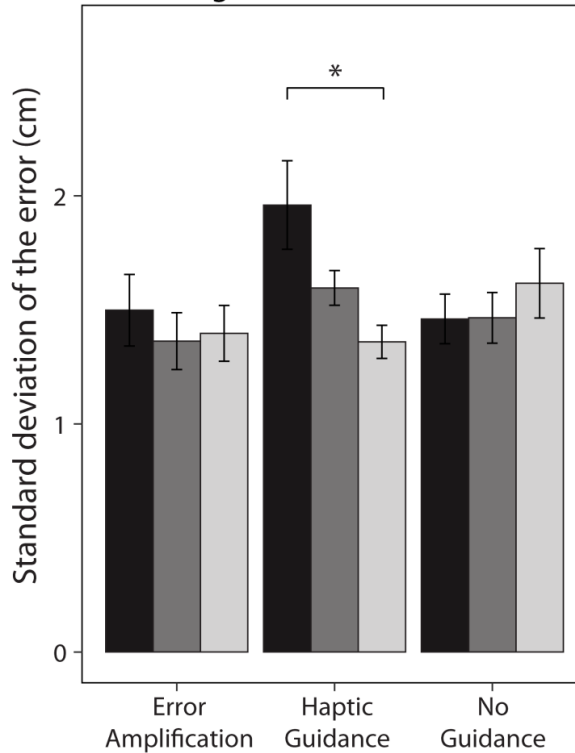


Fig. 7: Mean absolute error (A & B) and variability error (C & D) during the generalization tests (baseline, short-term and retention tests) for subjects trained with error amplification, haptic guidance and no guidance, during the continuous (wire) game (A & C) and the discrete (bar) game (B & D). Error bars:  $\pm 1$  SE. \*\*  $p < 0.001$ , \*  $p < 0.05$ .

### 3.5 Effect of training strategies on motivation

We did not find significant differences between training groups in any of the IMI subscales after baseline. Training with the different robotic strategies had a significant effect on self-reported motivation. We found a significant effect of the training strategy in interest/enjoyment during training (Fig. 8A,  $p = 0.008$ ). The main effect of training strategy did not reach significance after retention (Fig. 8A,  $p = 0.106$ ). In particular, subjects in the EA group reported a lower interest/enjoyment than subjects in the HG group (training:  $p = 0.030$ ; retention:  $p = 0.055$ ), and than subjects in the NG group (training:  $p = 0.002$ ; retention:  $p = 0.096$ ). The training strategy also influenced the perceived competence during training (Fig. 8B,  $p < 0.001$ ). In particular, subjects trained with EA reported a lower perceived competence level compared to subjects trained with NG ( $p = 0.001$ ) and HG ( $p < 0.001$ ). The training strategy had no significant effect on effort/importance and pressure/tension.

The game (continuous/wire vs. discrete/bar) also had a significant effect on motivation. Subjects reported greater values of interest/enjoyment in the bar game after baseline ( $p = 0.045$ ). Subjects also reported greater values of perceived competence in the bar game after baseline ( $p < 0.001$ ). The perceived competence in the wire game increased significantly more during training and retention than in the bar game (training:  $p = 0.038$ ; retention:  $p = 0.016$ ), especially in subjects trained with NG and HG. Subjects in the EA group, however, did not report differences between games. Subjects did not report differences in effort/importance between games after baseline. However, they reported greater values of effort/importance during training of the wire game ( $p = 0.016$ ). In fact, subjects increased the effort/importance during training of the wire game, but reduced it during training of the bar game ( $p = 0.049$ ). Subjects also reported greater values of pressure/tension in the wire game after baseline ( $p = 0.031$ ) and training ( $p = 0.039$ ).

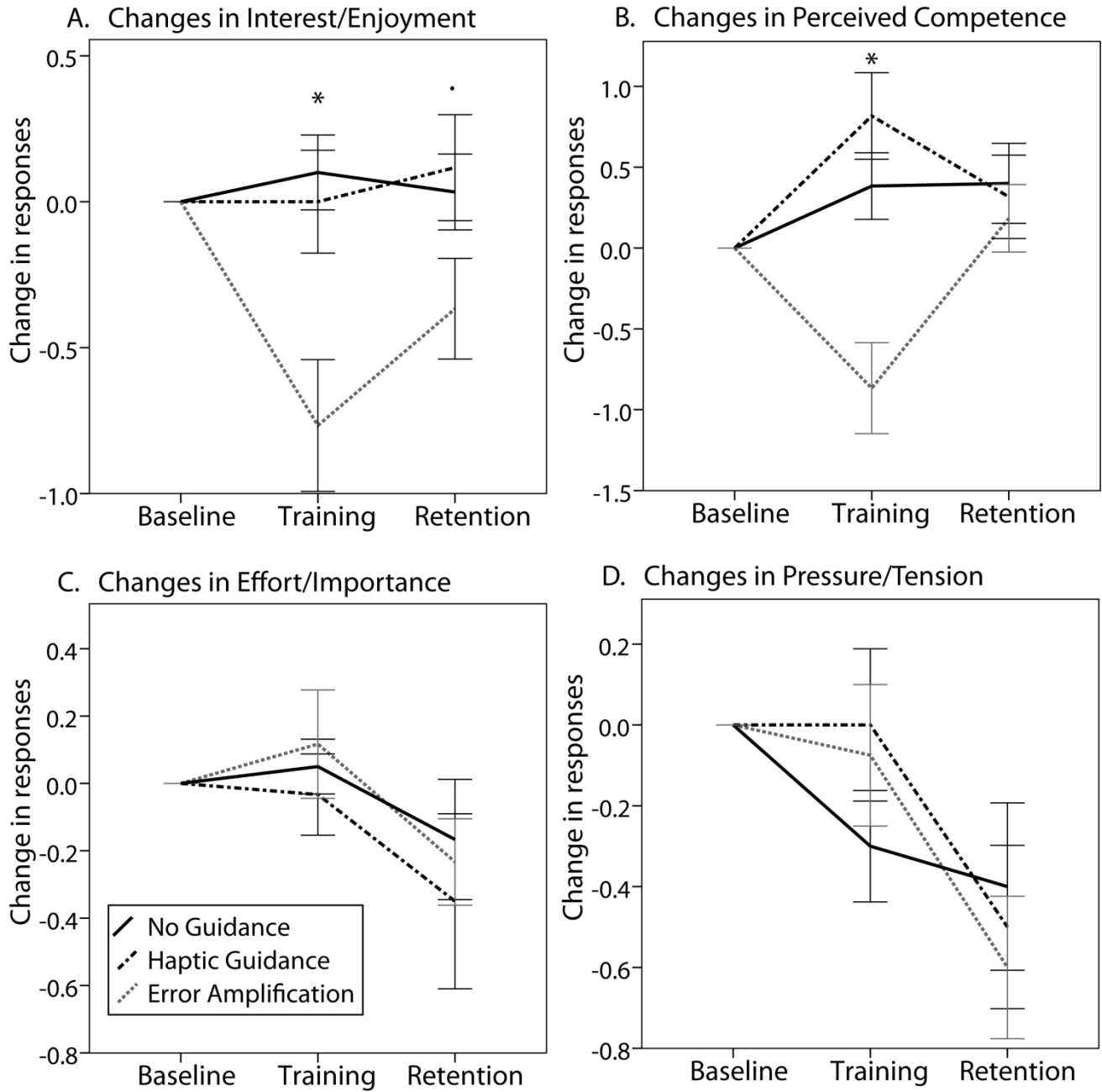


Fig. 8: Effect of the training strategies on self-reported motivation after the last training block and after retention test. **A:** Changes in responses to IMI questions related to interest/enjoyment. **B:** Changes in perceived competence. **C:** Changes in effort/importance. **D:** Changes in pressure/tension. All values are referred to the responses provided after baseline. \*  $p < 0.05$ ,  $\cdot p < 0.1$ . The asterisks show significant differences between training conditions for a given experimental phase. Error bars:  $\pm 1$  SE.

#### 4. Discussion

Previous research suggested that the effectiveness of robotic training strategies that modulate movement errors depend on subjects' initial skill level and the specific characteristics of the task to be learned (Marchal-Crespo et al. 2015). Motor tasks can be classified based on the time characteristics of the task, in particular the task's duration (discrete vs. continuous). Continuous tasks require movements without distinct beginning or end. Discrete tasks require fast movements that

include well-defined postures at the beginning and the end. In this paper, we evaluated the impact of three error-modulating training strategies –no guidance, error amplification and haptic guidance– on self-reported motivation and learning of two games, one that requires a continuous movement –a tracking task– and one that requires discrete movements –a fast reaching task. In order to rule out the effect of the initial skill level on learning, we successfully adapted the games difficulty to the subject's initial skill level.

All robotic strategies worked as expected during training, i.e., the haptic guidance controller improved subjects' performance, while the error amplification controller degraded it. As hypothesized, we found that training with error amplification resulted in better motor learning than haptic guidance. This is in line with previous studies that found that training with error amplification outperforms training with haptic guidance when the task difficulty is appropriate for subjects' initial skill level (Milot et al. 2010; Marchal-Crespo et al. 2014a; Duarte and Reinkensmeyer 2015; Marchal-Crespo et al. 2017). This can be explained by the Challenge Point Theory, which states that optimal learning is achieved when the difficulty of the task is appropriate for the individual subject's level of expertise (Guadagnoli and Lee 2004). Thus, adapting the game difficulty to the trainee's skill level provided the best framework for error amplification to enhance learning.

Only subjects trained with error amplification improved their performance after training the discrete task game, while training without guidance and with haptic guidance did not promote learning. In fact, subjects trained without guidance improved the performance in the discrete task game in a smaller amount than in the continuous task game. Interestingly, subjects in the no guidance group increased the interest/enjoyment in the continuous task game during training and retention, but reduced them in the discrete task game. Hence, it seems that amplifying errors during training benefits the process of learning in the discrete task game, whereas the other two training strategies hampered learning. In the continuous task game, however, we did not find a significant greater benefit of training with amplified error, compared to training with haptic guidance and no guidance. These divergent findings between games may originate from the different motor control mechanisms which are mainly needed to perform the games (feedback vs. feedforward motor control). Each brain area related to error-based control has its specific characteristics while performing a motor task that require feedback or feedforward motor control (Seidler et al. 2013). Therefore, it is possible that a training strategy that enhance learning of tasks that mainly require feedforward control (e.g. the bar game) did not enhance learning of tasks that require feedback motor control (e.g. the wire loop game).

Another possible rationale for the different findings in continuous and discrete motor tasks learning outcomes might originate from the different types of feedback that subjects received while training the games. Whilst in the continuous task game subjects continuously received visual feedback about their performance (i.e. concurrent visual feedback), in the discrete task game visual feedback was provided only once the target bar reached the ring horizontal position (i.e. terminal feedback). Hence, subjects experienced fewer errors while playing the discrete task game compared to the continuous task game. Research on motor learning has suggested that concurrent feedback attracts an external focus of attention during training (Shea and Wulf 1999), which was found to be beneficial for motor learning (Wulf 2013). In fact, subjects in this study reported greater values of effort/importance and pressure/tension during training of the feedback game, suggesting that the feedback game required greater attentional levels. Additionally, the specificity-of-learning hypothesis states that learning is

most effective when training is performed involving the most crucial sensory information source needed to perform the motor task in retention tests. In our case, in order to perform the continuous task game, the concurrent visual information was crucial, and therefore, maybe other sources of sensory information –for example proprioception– were neglected (Proteau 2005). Therefore, the concurrent experience of visual feedback, the increased attention and the maybe too long lasting training times in the continuous task game might have boosted the reduction of errors during training with haptic and no guidance strategies also, and not only after training with augmented errors. In fact, when looking at the reduction of the error variability –that was found to decrease with practice at a slower rate than mean error (Cohen and Sternad 2009)– after training the continuous task game, we found that subjects in the error amplification group tended to play more consistently than subjects in the haptic guidance group. This difference suggests that playing with amplified errors might be also beneficial to accelerate learning in the continuous task game.

In previous studies, we found divergent effects of training with error amplification on generalization. Training with error amplification enhanced generalization in a discrete task –e.g. playing a pinball-like game (Milot et al. 2010)– but limited transfer of learning in a continuous task –e.g. tracking a Lissajous figure (Marchal-Crespo et al. 2017). In the current study, all training strategies generalized learning of the continuous task game, but only subjects trained with haptic guidance generalized the continuous task game, although no significant differences between strategies were found. These contradictory results may derive from the substantial differences between the tasks that were used to test for generalization. In the previous studies, the generalization tasks were relatively similar to the trained tasks –e.g. extra untrained targets in the discrete task, and a slightly different Lissajous figure in the continuous task. In contrast, in this study the movement direction changed from vertical to horizontal, which is a considerable greater change than in previous studies.

Subjects in the error amplification group reported lower levels of interest and enjoyment during training than subjects trained with reduced errors (haptic guidance) and without robotic guidance, as hypothesized. Surprisingly, 3-5 days after training, subjects seemed to report lower values of interest and enjoyment, even whether they played without error amplification during the retention tests. We also found lower values of perceived competence during training with error amplification. However, this decreased feeling of competence did not persist at retention tests. This is in line with a recent study that found that subjects who trained with a robotic strategy that augmented kinematic variability when learning a golf putting task reported lower enjoyment levels that persisted after 1-3 days after training (Duarte and Reinkensmeyer 2015). However, contrary to this previous study, the reduced feelings of enjoyment and competence did not negatively affect the learning outcomes. We hypothesized an increase of reported effort during training with amplified errors, which could have reduced the negative effects of low enjoyment and perceived competence on motor learning outcomes. However, the differences between training groups in effort and importance did not reach significance. Maybe the relative short training times (subjects trained each game only during 4 minutes), limited the potential negative effects of low motivation in the acquisition of novel motor skills. Additionally, it is important to note that the responses related to effort do not necessary reflect the real subjects' mental or physical effort, but their subjective feelings regarding their effort. The error amplification strategy employed in this experiment differs from those previously presented in literature, which provide repulsive forces proportional to errors (Marchal-Crespo et al. 2014b; Duarte and Reinkensmeyer 2015), forces proportional to velocity and perpendicular to movement direction



(Patton et al. 2006), constant disturbing forces (Casellato et al. 2012), or randomly varying, unpredictable disturbing forces (Lee and Choi 2010). Maybe, the fact that error amplification was applied only in a small predefined area around the wire, and that the conversion towards haptic guidance prevented the ring to leave the wire, limited the subjects' subjective feeling of effort. Therefore, we cannot assume that the amount of physical effort required to fulfil the task with error amplification was, indeed, not greater than the effort required to perform the task with haptic guidance (Marchal-Crespo et al. 2014b).

Most of the significant differences observed between training strategies at short term retention were also observed at retention on the second day; however, these differences did not reach significance. Subjects seemed to forget part of what they learned during the 3-5 overly between experimental days, making the differences less obvious at retention. These lack of lasting significant differences might be due to the maybe too long time between experimental days –retention tests are usually performed after only 1-2 days (Heuer and Lüttgen 2014; Duarte and Reinkensmeyer 2015)– and due to the relative short training time. For example, in an important experiment in which participants learned to position the hand at a desired location, authors found similar results for short-term retention (measured 20 min after training) and retention (measured the next day) (Winstein et al. 1994). An interesting direction for future research is to more closely examine retention and generalization at different time points (e.g. after 1, 3 and 7 days) following the training techniques used here.

The experimental design suffers from some limitations. First, we ran a parallel design in order to reduce carryover effects that would affect learning (Marchal-Crespo et al. 2014b). However, a significant issue with parallel studies with a relative small number of sample points is bias (a systematic inaccuracy). In order to reduce bias, we randomized subjects into the different training strategies. We tested whether the randomization was enough to remove bias. We found that the mean errors during baseline 2 were not significantly different between groups, neither were the IMI responses related to each of the four IMI subscales. Secondly, the parameters employed in equations 1 and 2 to adjust the task difficulty based on subjects' performance were different between games. Therefore, we cannot assume that the two game adjustments were equivalent. In fact, subjects created significantly greater errors and reported lower perceived competence values when playing the wire game during baseline 2, compared to the bar game, suggesting that playing the bar game might have been easier. These differences might have also influenced the results. In fact, in previous experiments, we found that error-augmenting strategies were more suitable to enhance learning of especially simple tasks (Marchal-Crespo et al. 2014b). Finally, error amplifying strategies tend to alter the effective stiffness of the limbs through co-contraction of muscles (Franklin et al. 2007; Marchal-Crespo et al. 2014b). Therefore, in the discrete game, the observed error reduction after training with error amplification, but not in the haptic guidance and no guidance groups, might be due to a change of the arm stiffness. However, if stiffness would have been the solely reason for better performance, a better performance would have also been expected in the continuous task. Furthermore, co-contraction has been shown to fade rapidly once the strategy is removed (Marchal-Crespo et al. 2014b). An interesting direction for future research is to perform further studies while measuring upper arm muscle activity during training and retention.

Cortical changes have been shown to occur only with learning of new skills and not just with repetitive use (Lotze et al. 2003), suggesting that motor learning mechanisms are operative and

critical during brain plastic recovery. Therefore, understanding the underlying mechanisms of learning during the acquisition of novel motor skills may provide novel ways to improve neurorehabilitation therapies (Dietz and Ward 2015). Most of the motor learning studies and clinical studies with rehabilitation robots performed to date evaluated the effectiveness of robotic training strategies on motor learning and rehabilitation outcomes, independently of the motor task to be performed. However, tasks usually employed in robot-assisted rehabilitation are very diverse, from continuous tasks such as figure tracking (Dipietro et al. 2007; Nordin et al. 2014), to point to point reaching tasks (Krebs et al. 2003; Patton et al. 2006; Wolbrecht et al. 2008). Therefore, our findings might have important implications in the field of rehabilitation robotics. Error-amplifying training strategies have a great potential to provoke better motor learning in discrete motor tasks. However, the observed reduction of enjoyment and perceived competence, and the lack of greater effort during training with error amplification, might limit their applicability in neurorehabilitation. Motivation is a key factor to engage neurological patients into repetitive and intensive practice (Reinkensmeyer and Housman 2007), which in turn is crucial to enhance recovery. However, the impact of presented error-modulating training strategies on motor learning and motivation may differ in neurological patients. Therefore, additional studies with neurological patients are needed.

## **5. Conclusion**

We developed two games, one that requires a continuous movement –a tracking task– and one that requires discrete movements –a fast reaching task.. We conducted an experiment to evaluate the role of the motor task characteristics on the effectiveness of three error-modulating training strategies –no guidance, error amplification and haptic guidance– on self-reported motivation and learning of the continuous and discrete motor task games. Training with error amplification resulted in better motor learning than training with haptic guidance, besides the fact that error amplification reduced the interest/enjoyment and perceived competence during training. Only subjects trained with error amplification improved their performance after training the discrete task game, suggesting that error-amplifying training strategies have a great potential to provoke better motor learning, especially in tasks that require mainly feedforward motor control. It is, however, an open question whether the different findings in continuous and discrete motor task learning outcomes originate from the different nature of the motor control strategies mainly employed to fulfil the games, or whether differences arise from the different types of feedback that subjects received while training the games.

## **Conflict of Interest**

The authors declare that they have no conflict of interest.

## **Acknowledgments**

The authors gratefully acknowledge the contribution of Aniket Nagle, Anja Gut, Killian Baur, Fabian Just, Jaime Duarte, and Verena Klamroth-Marganska. The authors thank the Statistical Consulting service at ETH, Zurich, for their assistance in the statistical analysis. This work was partially supported by the Swiss National Science Foundation (SNF) through the grant number PMPDP2\_151319 and the National Centre of Competence in Research (NCCR) Robotics.

## **Ethical approval**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

Bates (2015) Fitting linear mixed-effects models using lme4.

Casellato C, Pedrocchi A, Zorzi G, et al (2012) Error-enhancing robot therapy to induce motor control improvement in childhood onset primary dystonia. *J NeuroEngineering Rehabil* 9:46. doi: 10.1186/1743-0003-9-46

Cohen RG, Sternad D (2009) Variability in motor learning: relocating, channeling and reducing noise. *Exp Brain Res* 193:69–83. doi: 10.1007/s00221-008-1596-1

Dietz V, Ward N (eds) (2015) The applicability of motor learning to neurorehabilitation. In: *Oxford Textbook of Neurorehabilitation*. Oxford University Press, pp 55–64

Dipietro L, Krebs HI, Fasoli SE, et al (2007) Changing Motor Synergies in Chronic Stroke. *J Neurophysiol* 98:757–768. doi: 10.1152/jn.01295.2006

Duarte JE, Reinkensmeyer DJ (2015) Effects of robotically modulating kinematic variability on motor skill learning and motivation. *J Neurophysiol* 113:2682–2691. doi: 10.1152/jn.00163.2014

Emken JL, Benitez R, Reinkensmeyer DJ (2007) Human-robot cooperative movement training: learning a novel sensory motor transformation during walking with robotic assistance-as-needed. *J Neuroengineering Rehabil* 4:8. doi: 10.1186/1743-0003-4-8

Emken JL, Reinkensmeyer DJ (2005) Robot-enhanced motor learning: accelerating internal model formation during locomotion by transient dynamic amplification. *IEEE Trans Neural Syst Rehabil Eng Publ IEEE Eng Med Biol Soc* 13:33–39. doi: 10.1109/TNSRE.2004.843173

Flash T, Hogan N (1985) The coordination of arm movements: an experimentally confirmed mathematical model. *J Neurosci* 5:1688–1703.

Franklin DW, Liaw G, Milner TE, et al (2007) Endpoint stiffness of the arm is directionally tuned to instability in the environment. *J Neurosci Off J Soc Neurosci* 27:7705–7716. doi: 10.1523/JNEUROSCI.0968-07.2007

Guadagnoli MA, Lee TD (2004) Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav* 36:212–224. doi: 10.3200/JMBR.36.2.212-224

Guidali M, Duschau-Wicke A, Broggi S, et al (2011) A robotic system to train activities of daily living in a virtual environment. *Med Biol Eng Comput* 49:1213–1223. doi: 10.1007/s11517-011-0809-0

Heuer H, Lüttgen J (2014) Motor learning with fading and growing haptic guidance. *Exp Brain Res* 232:2229–2242. doi: 10.1007/s00221-014-3914-0

- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J Biom Z* 50:346–363. doi: 10.1002/bimj.200810425
- Israel JF, Campbell DD, Kahn JH, Hornby TG (2006) Metabolic costs and muscle activity patterns during robotic- and therapist-assisted treadmill walking in individuals with incomplete spinal cord injury. *Phys Ther* 86:1466–1478. doi: 10.2522/ptj.20050266
- Kawato M (1999) Internal models for motor control and trajectory planning. *Curr Opin Neurobiol* 9:718–727.
- Kitago T, Krakauer JW (2013) Motor learning principles for neurorehabilitation. *Handb Clin Neurol* 110:93–103. doi: 10.1016/B978-0-444-52901-5.00008-3
- Klein J, Spencer S, Allington J, et al (2010) Optimization of a parallel shoulder mechanism to achieve a high-force, low-mass, robotic-arm exoskeleton. *Robot IEEE Trans On* 1–6.
- Krebs HI, Palazzolo JJ, Dipietro L, et al (2003) Rehabilitation Robotics: Performance-Based Progressive Robot-Assisted Therapy. *Auton Robots* 15:7–20. doi: 10.1023/A:1024494031121
- Lee J, Choi S (2010) Effects of haptic guidance and disturbance on motor learning: Potential advantage of haptic disturbance. In: 2010 IEEE Haptics Symposium. pp 335–342
- Lotze M, Braun C, Birbaumer N, et al (2003) Motor learning elicited by voluntary drive. *Brain J Neurol* 126:866–872.
- Marchal-Crespo L, Furumasu J, Reinkensmeyer DJ (2010a) A robotic wheelchair trainer: design overview and a feasibility study. *J NeuroEngineering Rehabil* 7:40. doi: 10.1186/1743-0003-7-40
- Marchal-Crespo L, López-Olóriz J, Jaeger L, Riener R (2014a) Optimizing learning of a locomotor task: amplifying errors as needed. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2014*:5304–5307. doi: 10.1109/EMBC.2014.6944823
- Marchal-Crespo L, McHughen S, Cramer SC, Reinkensmeyer DJ (2010b) The effect of haptic guidance, aging, and initial skill level on motor learning of a steering task. *Exp Brain Res* 201:209–220. doi: 10.1007/s00221-009-2026-8
- Marchal-Crespo L, Michels L, Jaeger L, et al (2017) Effect of error augmentation on brain activation and motor learning of a complex locomotor task. *Front Neurosci*. doi: 10.3389/fnins.2017.00526
- Marchal-Crespo L, Reinkensmeyer DJ (2009) Review of control strategies for robotic movement training after neurologic injury. *J NeuroEngineering Rehabil* 6:20. doi: 10.1186/1743-0003-6-20
- Marchal-Crespo L, Schneider J, Jaeger L, Riener R (2014b) Learning a locomotor task: with or without errors? *J Neuroengineering Rehabil* 11:25. doi: 10.1186/1743-0003-11-25

- Marchal-Crespo L, van Raaij M, Rauter G, et al (2013) The effect of haptic guidance and visual feedback on learning a complex tennis task. *Exp Brain Res* 231:277–291. doi: 10.1007/s00221-013-3690-2
- Marchal-Crespo L, Wolf P, Gerig N, et al (2015) The role of skill level and motor task characteristics on the effectiveness of robotic training: first results. In: 2015 IEEE International Conference on Rehabilitation Robotics (ICORR). pp 151–156
- McAuley E, Duncan T, Tammen VV (1989) Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. *Res Q Exerc Sport* 60:48–58. doi: 10.1080/02701367.1989.10607413
- Milot M-H, Marchal-Crespo L, Green CS, et al (2010) Comparison of error-amplification and haptic-guidance training techniques for learning of a timing-based motor task by healthy individuals. *Exp Brain Res* 201:119–131. doi: 10.1007/s00221-009-2014-z
- Muratori LM, Lamberg EM, Quinn L, Duff SV (2013) Applying principles of motor learning and control to upper extremity rehabilitation. *J Hand Ther Off J Am Soc Hand Ther* 26:94–102; quiz 103. doi: 10.1016/j.jht.2012.12.007
- Nef T, Guidali M, Riener R (2009) ARMin III &#x2013; Arm Therapy Exoskeleton with an Ergonomic Shoulder Actuation. *Appl Bionics Biomech* 6:127–142. doi: 10.1080/11762320902840179
- Nordin N, Xie SQ, Wünsche B (2014) Assessment of movement quality in robot- assisted upper limb rehabilitation after stroke: a review. *J Neuroengineering Rehabil* 11:137. doi: 10.1186/1743-0003-11-137
- Novak D, Nagle A, Keller U, Riener R (2014) Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. *J Neuroengineering Rehabil* 11:64. doi: 10.1186/1743-0003-11-64
- Patton JL, Stoykov ME, Kovic M, Mussa-Ivaldi FA (2006) Evaluation of robotic training forces that either enhance or reduce error in chronic hemiparetic stroke survivors. *Exp Brain Res* 168:368–383. doi: 10.1007/s00221-005-0097-8
- Proteau L (2005) Visual afferent information dominates other sources of afferent information during mixed practice of a video-aiming task. *Exp Brain Res* 161:441–456. doi: 10.1007/s00221-004-2090-z
- Reinkensmeyer DJ, Akoner O, Ferris DP, Gordon KE (2009) Slacking by the human motor system: computational models and implications for robotic orthoses. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2009:2129–2132. doi: 10.1109/IEMBS.2009.5333978
- Reinkensmeyer DJ, Housman SJ (2007) “If I can’t do it once, why do it a hundred times?”: Connecting volition to movement success in a virtual environment motivates people to exercise the arm after stroke. In: 2007 Virtual Rehabilitation. pp 44–48

- Richard Schmidt, Tim Lee (2010) *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL, USA: Human Kinetics Publishers, 2005.
- Riener R, Lünenburger L, Jezernik S, et al (2005) Patient-cooperative strategies for robot-aided treadmill training: first experimental results. *IEEE Trans Neural Syst Rehabil Eng Publ IEEE Eng Med Biol Soc* 13:380–394. doi: 10.1109/TNSRE.2005.848628
- Rüdt S, Moos M, Seppay S, et al (2016) Towards more efficient robotic gait training: A novel controller to modulate movement errors. In: 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob). pp 876–881
- Ryan RM (1982) Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *J Pers Soc Psychol* 43:450–461. doi: 10.1037/0022-3514.43.3.450
- Salmoni AW, Schmidt RA, Walter CB (1984) Knowledge of results and motor learning: a review and critical reappraisal. *Psychol Bull* 95:355–386.
- Scheidt RA, Reinkensmeyer DJ, Conditt MA, et al (2000) Persistence of motor adaptation during constrained, multi-joint, arm movements. *J Neurophysiol* 84:853–862.
- Schmidt RA, Bjork RA (1992) New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychol Sci* 3:207–217. doi: 10.1111/j.1467-9280.1992.tb00029.x
- Seidler RD, Kwak Y, Fling BW, Bernard JA (2013) Neurocognitive mechanisms of error-based motor learning. *Adv Exp Med Biol* 782:39–60. doi: 10.1007/978-1-4614-5465-6\_3
- Seidler RD, Noll DC, Thiers G (2004) Feedforward and feedback processes in motor control. *NeuroImage* 22:1775–1783. doi: 10.1016/j.neuroimage.2004.05.003
- Shadmehr R, Smith MA, Krakauer JW (2010) Error correction, sensory prediction, and adaptation in motor control. *Annu Rev Neurosci* 33:89–108. doi: 10.1146/annurev-neuro-060909-153135
- Shea CH, Wulf G (1999) Enhancing motor learning through external-focus instructions and feedback. *Hum Mov Sci* 18:553–571. doi: 10.1016/S0167-9457(99)00031-7
- Sigrist R, Rauter G, Marchal-Crespo L, et al (2015) Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Exp Brain Res* 233:909–925. doi: 10.1007/s00221-014-4167-7
- Tong C, Flanagan JR (2003) Task-specific internal models for kinematic transformations. *J Neurophysiol* 90:578–585. doi: 10.1152/jn.01087.2002
- Winsten CJ, Pohl PS, Lewthwaite R (1994) Effects of physical guidance and knowledge of results on motor learning: support for the guidance hypothesis. *Res Q Exerc Sport* 65:316–323. doi: 10.1080/02701367.1994.10607635
- Wolbrecht ET, Chan V, Reinkensmeyer DJ, Bobrow JE (2008) Optimizing compliant, model-based robotic assistance to promote neurorehabilitation. *IEEE Trans Neural Syst Rehabil Eng Publ IEEE Eng Med Biol Soc* 16:286–297. doi: 10.1109/TNSRE.2008.918389

- Wolpert DM, Ghahramani Z, Flanagan JR (2001) Perspectives and problems in motor learning. *Trends Cogn Sci* 5:487–494.
- Wolpert DM, Miall RC, Kawato M (1998) Internal models in the cerebellum. *Trends Cogn Sci* 2:338–347.
- Wu G, Siegler S, Allard P, et al (2002) ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. *J Biomech* 35:543–548. doi: 10.1016/S0021-9290(01)00222-6
- Wulf G (2013) Attentional focus and motor learning: a review of 15 years. *Int Rev Sport Exerc Psychol* 6:77–104. doi: 10.1080/1750984X.2012.723728